Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

97117154.1

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

Nigel Rollis

DEN HAAG, DEN
THE HAGUE, 22/04/99
LA HAYE, LE

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.:
Demande n°: 97117154.1

Anmeldetag:
Date of filing:
Date de dépôt: 02/10/97

Anmelder:
Applicant(s):
Demandeur(s):

Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.
Berlin
GERMANY
National Public Health Institute
00300 Helsinki
FINLAND

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:

Novel nucleic acid molecule encoding a (poly) peptide co-segregating in mutated form with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED)

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/[redacted]/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

VOSSIUS & PARTNER G
PATENTANWÄLTE
EUROPEAN PATENT ATTORNEYS
Siebertstr. 4 - 81675 München

Our Ref.: B 2911 EP

3

Novel nucleic acid molecule encoding a (poly)peptide co-segregating in mutated form with Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED)

The present invention relates to a nucleic acid molecule encoding a (poly)peptide co-segregating in mutated form with Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED). The present invention further relates to a nucleic acid molecule deviating by at least one mutation from the nucleic acid molecule described above wherein said mutation co-segregates with APECED and is an insertion, a deletion, a substitution and/or an inversion, and wherein said mutation further results in a loss or a gain of function of the (poly)peptide encoded by said mutated nucleic acid molecule. Furthermore, the present invention relates to a vector comprising the nucleic acid molecules described above and to a host transformed with said vector. In addition, the present invention relates to a process of recombinantly producing a (poly)peptide encoded by the nucleic acid molecules described above comprising culturing or raising said host and isolating said (poly)peptide from said culture or said host. The present invention further relates to the (poly)peptide encoded by said nucleic acid molecules or produced by the process described above. Additionally, the present invention relates to an antibody that specifically recognizes said (poly)peptides. Moreover, the present invention relates to a method for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutation in the wild-type nucleic acid molecule described above or a mutated form of the (poly)peptide encoded by said mutated nucleic acid molecule in an immuno-assay using the antibody described above.

Self tolerance and the ability to discriminate between self and non-self antigens are central to the immune response. Autoimmunity develops following a loss of self tolerance. There are several hypotheses which have been suggested, reflecting possible mechanisms leading to an autoimmune response: These hypotheses comprise:

- Presentation of sequestered self antigens: immunological tolerance is not established when molecules of the body are hidden from the lymphoreticular system (e.g. in the lens of the eye, in sperm or the heart). If the tissues are damaged, an autoimmune response can develop.

- Cross-reactivity: in the case when a self antigen and an exogenous antigen cross-react, the shared epitope is presented to the immune system with a different carrier, allowing T helper cells to confer a signal to B cells with antibody receptors recognizing the epitope.
- Modification of auto-antigens: a modification of an auto-antigen may arise and if different, this altered antigen could be recognized as foreign and trigger an immune response.
- Viral infections: auto-antibodies can sometimes arise following viral infections.
- Ectopic expression of HLA class II antigens: class II antigens have a restricted tissue distribution. The tissues affected in autoimmune diseases may express class II antigens inappropriately.
- Regulatory defects: (1) T cells sometimes recognize self-antigens but fail to co-operate with B cells due to peripheral tolerance exerted by suppressor T cells. A failure in this regulatory mechanism could result in autoimmunity. (2) Polyclonal B cell activation: some molecules can mimic the T cell stimulus and activate B cells to divide polyclonally. This could lead to the activation of B cells secreting auto-antibodies.

There is a wide range of autoimmune diseases. The spectrum spans conditions involving a single organ through those involving all systems in the body. Autoimmune diseases are characterized by an abnormal response of the human immune system to self components. The impact of these diseases on health of populations is high since many common diseases like diabetes mellitus, multiple sclerosis or rheumatoid arthritis represent autoimmune reactions. Censequently, characterization of molecules involved in autoimmunity are of high importance for the cure and treatment of these disorders.

Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED, OMIM 240300) is an autosomal recessive disease characterized by 1) autoimmune polyendocrinopathies: hypoparathyroidism, adrenocortical failure, IDDM, gonadal failure, hypothyroidism, pernicious anemia, and hepatitis, 2) chronic mucocutaneous candidiasis and 3) ectodermal dystrophies: vitiligo, alopecia, keratopathy, dystrophy of dental enamel, nails and tympanic membranes (Ahonen, P., et al., *N. Engl. J. Med.*, **322**, 1829-1836 (1990)). The disease is reported worldwide but is exceptionally prevalent among the Finnish population (incidence 1: 25 000) and the Iranian Jews (Ahonen, P., et al., *N. Engl. J. Med.*, **322**, 1829-

1836 (1990); Zlotogora, J., et al., *J. Med. Genet.*, **29**, 824-826 (1992)). The primary biochemical defect in this disorder remains elusive.

APECED is the only described systemic autoimmune disease in humans with Mendelian inheritance, and the clinical phenotype characterized by autoimmune endocrinopathies, including IDDM, and chronic candidiasis would suggest defects in both humoral (Ahonen, P., et al., *J. Clin. Endocrinology and Metabolism*, **64**, 494-500 (1987)) and cell mediated immunity (Fidel, P. L. & Sobel, J. D., *TIMB*, **2**, 202-206 (1994)). No single HLA associated haplotype exists (Ahonen, P., et al., *J. Clin. Endocrinology and Metabolism*, **66**, 1152-1157 (1988)), autoantibodies are found against several cell types in the patients' sera (Ahonen, P., et al., *J. Clin. Endocrinology and Metabolism*, **64**, 494-500 (1987)) and only unspecific abnormal responses have been found in T cell proliferation tests. These observations would suggest a deregulation of both B and T cell specific immune responses in APECED. Moreover, the non-specific autoantibodies detected in the APECED patients' sera against several cell types do not support the hypothesis of one major autoantigen (Krohn, K., et al., *Lancet*, **339**, 770-773 (1992)). However, despite these well defined characteristics, the etiology of APECED, like of most autoimmune diseases, remains unknown. Insights into said etiology would also provide an entry point for the dissection of molecular mechanisms leading to the development of autoimmunity in general. On the basis of such knowledge, means and methods for the prevention or treatment of autoimmune diseases in general and APECED in particular might be developed.

Accordingly, the technical problem underlying the present invention was to uncover factors involved in the development of APECED that might contribute to providing means of treating or curing monogenic autoimmune diseases, in particular APECED.

The solution to the above technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, in one aspect the present invention relates to a nucleic acid molecule encoding a (poly)peptide co-segregating in mutated form with Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED) which is

(a) a nucleic acid molecule comprising a nucleic acid molecule encoding the (poly)peptide having the amino acid sequence of Fig. 2A;

(b) a nucleic acid molecule comprising the nucleic acid molecule having the nucleotide sequence of Fig. 2A that encodes the amino acid sequence of Fig. 2A;

(c) a nucleic acid molecule hybridizing to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule which is degenerate to the nucleic acid molecule of (c).

The present invention surprisingly revealed that a novel polypeptide, designated APGD1 for autoimmune polyglandular disease type 1, encoded by the nucleic acid molecule of the invention co-segregates in mutated form with APECED.

As used herein, the term "co-segregation" relates to any association of the mutated form of the polypeptide with APECED. APGD1 is a protein with a predicted length of 545 amino acids, a theoretical molecular weight of 57,7 kD and a calculated pI of 7,53. Statistical analysis of the protein sequence of Fig. 2A (Brendel, V., et al., *Proc. Natl. Acad. Sci. USA*, **89**, 2002-2006 (1992)) indicates a high content of proline (11.7%) but no apparent clusters of charged amino acids or periodicity patterns. The secondary structural content of APGD1 was predicted to consist mostly of coils, with only a weak probability for the occurrence of structural α-helixes or β-sheets. A putative bi-partite nuclear targeting signal (Dingwall, C. & Laskey R. A.,*TIBS*, **16**, 478-481 (1991)) was found between amino acids 113 to 133 (Figure 2A). The predicted protein harbors two cysteine-rich regions of 42 amino acids, each specifying a Cys4-His-Cys3 double-paired finger motif similar to the PHD finger type (Aasland, R., et al., *TIBS*, **20**, 56-59 (1995)) (Figure 2A). Spacing of essential residues is conserved in the two motifs found in APGD1: $C_{299,434}$-XX-$C_{302,437}$-X(8)-$C_{311,446}$-XX-$C_{314,449}$-X(4)-$H_{319,454}$-XX-$C_{322,457}$-X(14)-$C_{337,471}$-XX-$C_{340,474}$ (where X is any amino acid and numbers in parenthesis represent the length of the intervening peptide sequence). This structural motif has been reported for a number of nuclear proteins involved in the mediation or regulation of transcription, such as TIF1 (Transcription Intermediary Factor 1) (Douarin, Le, B., et al., *EMBO J.*, **14**, 2020-2033 (1995)) and KRIP-1 (KRAB-A Interacting Protein) (Kim, S-S., et al., *Proc. Natl. Acad. Sci, USA*, **13**, 15299-14304 (1996)). Sequence homology of APGD1 with other proteins in the databases was strictly limited to this Cys4-His-Cys3 motif. Although the spacing of residues is conserved in each case, the sequence is most closely homologous to the Mi-2 autoantigen (Ge, Q., et al., *J. Clin.*

Invest., **96**, 1730-1737 (1995)) and the TIF1 proteins (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)). Mi-2 is the major nuclear antigen detected in the sera of autoimmune dermatomyositis patients (Ge, Q., et al., *J. Clin. Invest.*, **96**, 1730-1737 (1995)) and TIF1 is involved in the transcriptional control of the estrogen receptor (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)).

By the provision of the nucleotide acid molecule of the invention it is now possible to isolate identical or similar nucleic acid molecules which code for proteins with identical functions and characteristics and which are derived from other individuals or which represent alleles of the nucleic acid molecule of the invention. Well-established approaches for the identification and isolation of such related sequences are, e.g., the isolation from genomic or cDNA libraries using the complete part of the disclosed sequence as a probe or the amplification of corresponding nucleic acid molecules by polymerase chain reaction using specific primers.

As stated hereinabove, the invention also relates to nucleic acid molecules which hybridize to the above described nucleic acid molecules and differ at one or more positions in comparison to these as long as they encode a (poly)peptide having the above described characteristics. In connection with the present invention, the term "hybridizing" is understood as referring to conventional hybridization conditions, preferably such as hybridization in 50% formamide, 6x SSC, 0.1% SDS, and 100µg/ml ssDNA, in which temperatures for hybridization are above 37°C and temperatures for washing in 0.1x SSC, 0.1% SDS are above 55°C. Most preferably, the term "hybridizing" refers to stringent hybridization conditions, for example such as described in Sambrook, et al. (Molecular cloning; A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1989)) or Higgins & Hames (Nucleic acid hybridization, A practical approach, IRL Press, Oxford (1985)). Said nucleic acid molecules comprise those which differ, for example, by deletion(s), insertion(s), alteration(s) or any other modification known in the art in comparison to the above described nucleic acid molecules. Methods for introducing such modifications in the nucleic acid molecules according to the invention are well-known to the person skilled in the art; see, e.g., Sambrook, et al., supra.

As mentioned hereinabove, the invention also relates to nucleic acid molecules the sequence of which differs from the sequence of the above-described hybridizing molecules due to the degeneracy of the genetic code.

In a preferred embodiment of the nucleic acid molecule of the present invention, said (poly)peptide has the function of a transcription factor or a transcription-associated factor. As used herein, the term "transcription factor" or "transcription-associated factor" comprises any factor which directly or indirectly influences transcription of a gene by, e.g., directly interacting with regulatory sequences, interacting with other transcription regulating factors, changing the conformation of chromatin, and the like.

The (poly)peptide encoded by the nucleic acid molecule of the invention preferably comprises at least one zinc finger motif. The term "zinc finger" describes a certain amino acid motif, which is able to bind metal ions, and is well known for those skilled in the art. Preferably, the (poly)peptide of the invention comprises two double-paired zinc finger motifs.

In a further embodiment, the present invention relates to a nucleic acid molecule deviating by at least one mutation from the nucleic acid molecules described above, wherein said mutation co-segregates with APECED and is

(a) an insertion;
(b) a deletion;
(c) a substitution; and/or
(d) an inversion,

and wherein said mutation further results in a loss of function or a gain of function of the (poly)peptide of the invention.
The term "substitution", as used herein, also includes point mutations resulting in an amino acid exchange.

In a preferred embodiment of the present invention, said insertion, which is a duplication of 4 nucleotides (CCTG) normally found at position 1086-1089, is a 4 nucleotide insertion at

the nucleotide position 1085 or 1090 or an insertion of an adenosine at position 1284 of the nucleotide sequence of Fig. 2A.

In another preferred embodiment of the invention, said deletion is a 13 nucleotide deletion of nucleotides 1085 -1097 or a deletion of the cytosine at position 1313 of the nucleotide sequence of Fig. 2A .

In still another preferred embodiment of the present invention, said substitution is a cytosine to thymidine exchange at nucleotide position 889 of the nucleotide sequence of Fig. 2A.

As mentioned above, said mutation results in a loss or a gain of function of the (poly)peptide of the invention. In a preferred embodiment of the present invention, said loss of function is a loss of macromolecule binding properties. However, a loss of transactivating property in addition or instead of the loss of the macromolecule binding property is also envisaged. Other possibilities relate to the loss of a structural determinant (truncated protein) in addition to the loss of a functional determinant.

In an alternative preferred embodiment of the present invention, said gain of function is involved in molecular interaction. An example of such a gain of function is the indirect regulation of a cellular process. For instance, if the deletion of a zinc finger results in the loss of a binding property involving a second molecule, this second molecule may "gain" a function in case its function was modulated by APGD1.

The present invention further relates to a fragment of any of the aforementioned nucleic acid molecule(s) comprising at least 14 nucleotides. Preferably, said fragment is 17 nucleotides long, and most preferably, it is 21 nucleotides long. Said fragment can be used, e.g., as a probe in nucleic acid hybridization experiments like, e.g., Southern or Northern blot experiments, or as primer in primer extension analyses. In a preferred embodiment said fragment is labeled.

In another aspect, the present invention provides a nucleic acid molecule which is complementary to any of the nucleic acid molecules or fragments thereof described above. Such a nucleic acid molecule can be used, e.g., as a probe in RNase protection assays, or as an anti-sense probe to inhibit expression of the (poly)peptide(s) of the present invention. The person skilled in the art is familiar with the preparation and the use of said probes (see, e.g., Sambrook et al., supra).

In yet another aspect, the present invention relates to a mammalian homologue of the nucleic acid molecule(s) of the present invention. The person skilled in the art knows on the basis of the teachings of the present invention how to obtain the homologue, e.g., of other mammals such as mouse, rat, rabbit or pig. This can be effected, e.g., by hybridization of the molecule of the present invention under low stringent conditions to the corresponding nucleic acids from other species contained, e.g., in conventional libraries. "Low stringent conditions" differ from stringent conditions (described hereinabove) in that higher salt concentrations and/or lower temperatures are employed for hybridization. Such conditions are well known in the art (see, e.g., Sambrook et al. or Higgins & Hames, supra).

In a further embodiment of the present invention, the nucleic acid molecule(s) of the invention are DNA molecules like, e.g., cDNA or genomic DNA molecules, or RNA molecules like mRNA molecules.

In another embodiment, the present invention provides a primer pair which hybridizes under stringent conditions to any of the nucleic acid molecules mentioned above. Said primer pair can be used, e.g., in a polymerase chain reaction (PCR) to amplify nucleic acid fragments derived from the nucleic acid molecules described above. In the case that RNA is used as the template in the amplification reaction, it is beforehand reverse transcribed into DNA. The skilled artisan knows how to design and use said primer pair, which conditions for the amplification reaction have to be set up, and how to reverse transcribe RNA into DNA (see, e.g., Sambrook et al., supra).

Furthermore, the present invention relates to a vector comprising a nucleic acid molecule of the invention.

Examples for such vectors are pUC18/19, pBR322 or pBlueScript all of which are commercially available. In a preferred embodiment the nucleic acid molecule present in the vector is operatively linked to regulatory elements permitting expression in prokaryotic or eukaryotic host cells. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the *lac* or *trp* promoter in E. coli, and examples for regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40- , RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. Beside elements which are responsible for the initiation of transcription such regulatory elements may also comprise transcription termination signals, such as the SV40-poly-A site or the tk-poly-A site, downstream of the nucleic acid molecule of the invention.

The invention also relates to a host comprising a vector according to the invention. The transformation of hosts with the vectors of the invention is well known in the art (see, e.g., Sambrook et al., supra)

In a preferred embodiment of the present invention, the host is a bacterium, a yeast cell, an insect cell, a fungal cell, a mammalian cell, a plant cell, a transgenic animal or a transgenic plant. As used herein, the term "transgenic" also relates to organisms that contain a gene which has been knocked out. For example, animals with no functional allele of the APGD1-gene can be used for the investigation of the role APGD-1 plays in cellular life as well as a model for the development of APECED. Techniques for the production of transgenic or knock-out organisms are well known in the art.

In a further embodiment, the present invention relates to a process of producing a (poly)peptide of the invention comprising culturing or raising the host described above and isolating said (poly)peptide from said culture or said host. Such methods are well known in the art (see, e.g., Sambrook et al., supra).

Furthermore, the invention relates to a (poly)peptide encoded by a nucleic acid molecule of the invention or produced by the above described process. In this context it is also

understood that the (poly)peptides according to the invention may be further modified by conventional methods known in the art. By providing the (poly)peptides according to the present invention it is also possible to determine the portions relevant for their biological activity. This may allow the construction of chimeric proteins or fusion proteins comprising an amino acid sequence derived from a (poly)peptide of the invention which is crucial for its biological activity and other functional amino acid sequences like, e.g., nuclear localization signals, transactivating domains, DNA-binding domains, hormone-binding domains, protein tags (GST, GFP, h-myc peptide, Flag, HA peptide) which may be derived from the same or from heterologous proteins. Said chimeric or fusion proteins are also comprised by the present invention.

The present invention also relates to a compound derived from a (poly)peptide of the invention and having essentially the same three dimensional structure thereof. Said compounds can be theoretically constructed on computers using molecular modelling software and subsequently be synthesized. Since such compounds are preferably not of proteinaceous nature, they may be used in applications where proteolytic degradation should be avoided, e.g., when contained in pharmaceutical compositions that are applied orally. The design of such compounds may, e.g., be effected by peptidomimetics.

In a further embodiment, the present invention relates to an antibody that specifically recognizes the (poly)peptide of the invention. Namely, the invention relates to an antibody which specifically recognizes (poly)peptides according to the invention irrespective of whether they are the wild-type or a mutated form and/or depending on whether the (poly)peptide of the invention is the wild-type or a mutated form. The antibody of the present invention may be a monoclonal antibody, a polyclonal antibody or a synthetic antibody as well as a fragment of said antibodies, such as, e.g., a Fab, a Fv or a scFv fragment. Furthermore, the antibody or fragments thereof can be obtained by using methods which are described, e.g., in Harlow and Lane, "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. The antibody of the present invention can be used, e.g., for the immunoprecipitation and immunolocalization of the (poly)peptides of the invention as well as for the monitoring of the presence of such (poly)peptides, e.g., in recombinant organisms, and

for the identification of compounds interacting with the (poly)peptides according to the invention.

Moreover, the present invention relates to a pharmaceutical composition comprising at least one of the aforementioned nucleic acid molecules, vectors, (poly)peptides, three-dimensionally equivalent compounds, and/or the antibody according to the present invention either alone or in combination, and optionally a pharmaceutically acceptable carrier. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by conventional methods. The pharmaceutical compositions can be administered to the subject at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g. by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Generally, the regimen as a regular administration of the pharmaceutical composition should preferably be in the range of 1 µg to 10 mg units per day. If the regimen is a continuous infusion, it should preferably also be in the range of 1 µg to 10 mg units per kilogram of body weight per minute, respectively. Progress can be monitored by periodic assessment. Dosages will vary but a preferred dosage for intravenous administration of DNA is preferably from approximately 10^6 to 10^{22} copies of the DNA molecule. The compositions of the invention may be administered locally or systemically. Administration will generally be parenterally, e.g., intravenously; DNA may also be administered directly to the target site, e.g., by biolistic delivery to an internal or external target site or by catheter to a site in an artery.

In addition, the present invention relates to a diagnostic composition comprising at least one of the aforementioned nucleic acid molecules, vectors, (poly)peptides, three-

dimensionally equivalent compounds, and/or the antibody according to the present invention either alone or in combination.

Said diagnostic composition can be used to test for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutation in the nucleic acid molecule(s) of the invention. Furthermore, the diagnostic composition can be used to test for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutated form of the (poly)peptide(s) according to the invention in an immuno-assay using the antibody of the invention. The term "immuno-assay", as used herein, comprises methods like, e.g., immuno-precipitation, immuno-blotting, ELISA, RIA, indirect immuno-fluorescence experiments, and the like. Such techniques are well known in the art and are described, e.g. in Harlow and Lane, supra.

In another embodiment, the present invention relates to methods for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutation in the nucleic acid molecule(s) of the invention. Such methods comprise, e.g., amplifying nucleic acid molecules from a nucleic acid obtained from a prospective patient or from a person suspected of carrying a predisposition for APECED with the primer pair of the invention, and analyzing the amplified nucleic acid molecules for the presence of a mutation. Said nucleic acid molecules can be analyzed, e.g., by sequencing with the primer or probe of the invention, hybridizing with the primer of the invention or by size-fractionating said nucleic acid molecules by gel-electrophoresis. Alternatively, and by way of example said nucleic acid obtained from a prospective patient or from a person suspected of carrying a predisposition for APECED can be directly analyzed by sequencing or hybridizing with the primer or probe of the invention. All the above mentioned primers or probes may hybridize to a mutated or a wild-type sequence. Further, all of the aforedescribed methods are well known in the art (see, e.g., Sambrook et al., supra).

In yet another embodiment, the present invention relates to methods for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutated form of the (poly)peptide(s) according to the invention. Such methods comprise, e.g., immuno-precipitation, immuno-blotting, ELISA, RIA, indirect immuno-fluorescence experiments, and the like. Such techniques are well known in the art and are described, e.g. in Harlow and Lane, supra.

In another embodiment, the present invention relates to the use of the nucleic acid molecule(s) or the vectors of the invention for gene therapy. Vectors comprising a nucleic acid molecule of the invention may be stably integrated into the genome of the cell or may be maintained in an extrachromosomal form. On the other hand, viral vectors described in the prior art may be used for transfecting certain cells, tissues or organs. Suitable gene delivery systems may include liposomes, receptor-mediated delivery systems, naked DNA, and viral vectors such as herpes viruses, retroviruses, adenoviruses, and adeno-associated viruses, among others. Delivery of nucleic acid molecules to a specific site in the body for gene therapy may also be accomplished using biolistic delivery systems.

Standard methods for transfecting cells with nucleic acid molecules are well known to those skilled in the art, see, e.g., Sambrook et al., supra. Gene therapy to cure APECED may be carried out by directly administering the nucleic acid molecule of the invention encoding a functional form of APGD1 to a patient or by transfecting cells with said nucleic acid molecule of the invention ex vivo and infusing the transfected cells into the patient. Furthermore, research pertaining to gene transfer into cells of the germ line is one of the fastest growing fields in reproductive biology. Gene therapy, which is based on introducing therapeutic genes into cells by ex-vivo or in-vivo techniques is one of the most important applications of gene transfer. Suitable vectors and methods for in-vitro or in-vivo gene therapy are described in the literature and are known to the person skilled in the art. The nucleic acid molecules comprised in the pharmaceutical composition of the invention may be designed for direct introduction or for introduction via liposomes, or viral vectors (e.g. adenoviral, retroviral) containing said nucleic acid molecule into the cell. Preferably, said

cell is a germ line cell, embryonic cell, or egg cell or a cell derived therefrom, if the production of transgenic non-human animals is envisaged.

It is to be understood that the introduced nucleic acid molecule encoding the protein having the biological activity of APGD1 expresses said protein after introduction into said cell and preferably remains in this status during the lifetime of said cell. For example, cell lines which stably express said protein having the biological activity of APGD1 may be engineered according to methods well known to those skilled in the art. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with the recombinant DNA molecule or vector of the invention and a selectable marker, either on the same or separate vectors. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows for the selection of cells having stably integrated the plasmid into their chromosomes and growing to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the protein having the biological activity of APGD1. A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase, hypoxanthine-guanine phosphoribosyltransferase, and adenine phosphoribosyl-transferase in tk, hgprt or aprt cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate, gpt, which confers resistance to mycophenolic acid, neo, which confers resistance to the aminoglycoside G-418, hygro, which confers resistance to hygromycin, or puromycin (pat, puromycin N-acetyl transferase). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine, and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO.

The figures show:

Figure 1

A) The physical map of the APECED region showing the markers used to construct the disease haplotypes (cen - JA1, D21S1912, PFKL (CA_n), PB1, D21S171 - tel), the other genes (PFKL, green and 694N10, pink) and the ESTs (EST cluster 1: AA082879, AA085392, EST cluster 2: N67176, T84071, T86112, T79577, T79655, R23544, R44295, EST cluster 3: AA453553) located in the close vicinity of APGD1 (blue) and the key cosmid clones Q21D11 and Q22G11 used for genomic sequencing as well as cosmid clone Q11D11 that was used as orientation marker in the fiber FISH experiment (see Figure 1C).

B) The genomic structure of the APGD1 gene. The 14 true exons of the gene are compared with the gene models predicted with different gene finding programs (Uberbacher, E., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 11261-11265 (1991); Thomas, A., & Skolnick, M. H., *IMA J. Math. Appl. Med. Biol.*, **11**, 149-160 (1994); Kulp, D., et al., ISMB-96, St. Louis, MO, AAAI/MIT Press. (http://www-hgc.lbl.gov/projects/genie.html) (1996)). Solid boxes indicate exons in which at least one boundary was correctly predicted, open boxes are false exons. Genomic sequence of cosmid clones Q21D1, Q22G11, EST matches, detailed gene prediction data and the intron-exon boundaries of APGD1 are available at http://chr21.rz-berlin.mpg.de/APECED.html/.

C) Fiber FISH image showing the assignment of the APGD1, red signal, (cDNA clone B1-1 used as a probe) in relation to previously mapped cosmid clones, Q11D11 (yellow) and Q21D1 (green). Detailed protocol is described elsewhere (Heiskanen, M., et al., *TIG*, **10**, 379-382 (1996)).

Figure 2

A) The nucleotide and predicted amino acid sequence of APGD1. The boundaries corresponding to the composite cDNA sequence are indicated by brackets, the most 3' end nucleotides for cDNA clones B1-1 and D1-1 are at positions 1809 and 2181, respectively. The last 64 nucleotides were determined by PCR extension. A putative non-canonical polyadenylation signal was found at nucleotide 2191 (underlined). The Alu sequence overlapping with the PFKL promotor is starting at nucleotide 1995 (arrowed bracket). Silent polymorphisms are indicated by small arrows (nucleotides 708, 801, 1317 and 1698). The

predicted protein is 545 amino acids. The putative bi-partite nuclear localisation signal is underlined in blue. The two PHD zinc finger domains are underlined in magenta. The cDNA sequence has been deposited in EMBL (Accession No. Z97990).

B) Northern blot analysis using cDNA B1-1 (1,8 kb) as a probe on a multiple tissue Northern blot, each lane containing 2 µg poly(A) RNA from human adult tissues (Clontech catalog # 7754-1 and 7751-1). The lower panel shows the hybridization with the β-globin control probe.

Figure 3

The mutations in the APGD1 gene (see also Table 1). A) The C-lanes of the sequencing gel showing a patient homozygous for the Finnish major mutation and a normal control. C_{889} of the patient has been mutated to T. B) A-lanes of a normal control and a Finnish patient heterozygous for the haplotype 4.1 show an A insertion at position 1284. C) Homozygous deletion of C_{1313} is observed in C-lane of the sequence of a French patient also homozygous for the disease haplotype 5.1. D) Comparison of C-lanes of an Italian patient homozygous for the haplotype 2.1 and normal control reveal a 4 bp insertion (nucleotides 1086-1089). E) A 13 bp deletion (nucleotides 1085-1097) can be observed in C-lanes of a patient carrying haplotype 3.1 compared with a normal control.

The examples illustrate the invention

Example 1: Isolation of the APGD1-cDNA

We have mapped APECED to chromosome 21q22.3 by linkage analysis and further refined the localisation by linkage disequilibrium to a region between the markers D21S25 and D21S171 (Aaltonen, J., et al., *Nature Genet*, **8**, 83-87 (1994); Aaltonen et al., Genome Research 7 (1997), 820-827). This critical region was 350 kb in size and a bacterial clone contig was constructed across this region. Several techniques were used to identify candidate genes in this gene rich region. Exon trapping (Buckler, A., et al., *Proc. Natl. Acad. Sci.*, USA, **88**, 4005-4009, (1991)) and cDNA selection (Lovett, M., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 9628-9632, (1991)) methods identified a new gene, 694N10 (Accession No. Z93322), just distal to the previously known PFKL gene (Phosphofructokinase of liver type, EC 2.7.1.11) (Elson et al., *Genomics*, 7, 47-56 (1990)) (Figure 1A). Partial unordered genomic sequence encompassing the PFKL gene (available at the International Chromosome 21 genomic sequence

repository, http:// www-eri.uchsc. edu/chr21/ eridna.html) was used to generate a new polymorphic marker, PB1. This marker showed an obligatory recombination in one APECED family, thus we were able to restrict the APECED region to 145 kb between the markers D21S25 and PB1 (Figure 1A). Therefore 694N10 was excluded as causative gene for APECED.

In parallel, we initiated a large scale sequencing approach from cosmid clones 21D1 and 22G11 mapping to the critical region (Figure 1A). A total of 87 kb of genomic sequence obtained from these cosmids were analysed with BlastN and BlastX algorithms (Altschul, S. F., et al., *J. Mol. Biol.*, **215**, 403-410, (1990)) against public databases. Three different EST (Expressed Sequence Tag) clusters were found in a region between D21S25 and PFKL (Figure 1A). Exon prediction was performed using the GRAIL2 program (Uberbacher, E., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 11261-11265 (1991)). A gene model was predicted directly upstream of the promotor of PFKL where no EST matches were identified (exons G1 to G7, Figure 1B). However, since the linkage disequilibrium data (Björses, P., et al., *Am. J. Hum. Genet.*, **59**, 8779-886 (1996)) suggested the APECED gene to be located in the close vicinity of PFKL further analyses were focused on this potential gene. Polymerase Chain Reaction (PCR) amplification (5'- AGA AGT GCA TCC AGG TTG GC-3' and 5'-GGA AGA GGG GCG TCA GCA AT-3') of a 316 bp genomic fragment spanning predicted exons G5 and G6 (Figure 1B) generated a probe for screening a human adult thymus cDNA library (Clontech catalog # HL5010b). Two cDNA clones (B1-1 and D1-1) and a 3' UTR extension PCR product yielded a composite cDNA sequence of 2,245 kb (Figure 2A). The cDNA clone B1-1 was localised on the physical map by fiber FISH (Fluorescent In Situ Hybridization) (Figure 1C) (Heiskanen, M., et al., *TIG*, **10**, 379-382 (1996)). Northern blot analysis showed a major transcript of approximately 2 kb expressed in all tissues analysed, the most intensive signals were obtained from thymus, pancreas and adrenal cortex (Figure 2B). In this respect, it is surprising that no ESTs were found in the databases. The cDNA sequence exhibits an unusually high GC content of 68.8% and contains an open reading frame (ORF) of 581 amino acids followed by a STOP codon at nucleotide 1756. The likely initiator ATG codon occurs at nucleotide 121 (Figure 2A), predicting a 545 residue protein.

Example 2: Structure of the APGD1-gene

The structure of the APGD1 gene was determined from a comparison of the cDNA sequence with the cosmid 22G11 genomic sequence using the est_genome program (developed by Richard Mott, available at the Sanger center, UK). The genomic structure consists of 14 exons spanning 11,9 kb of genomic DNA (Figure 1B). A putative promotor containing a TATA box located 35 nucleotides from the first nucleotide of exon 1 and a GC box was identified immediately upstream of the first exon of the APGD1 gene. A CpG island was also associated with the promotor region. Detailed analysis of the genomic sequence upstream of the APGD1 gene did not suggest any additional exons within 22 kb of the predicted promotor. The translation of the genomic sequence identified an in frame STOP codon 16 residues upstream of the first amino acid of the translated cDNA sequence. Analysis of the 3' end of the gene suggested that exon 14 represents the last exon since the STOP codon at position 1756 is followed by repetitive sequences. Further, exon 14 overlaps with the promoter region of the PFKL gene (Levanon, D., et al., *Biochem and Mol. Biol. Int.*, **35**, 929-936 (1995)) which is transcribed from the same DNA strand (Figure 1B and 2A). Apparent C to T silent polymorphisms were found at third codon positions in exons 5, 6, 10 and 14 (Figure 2A). The gene organisation was poorly predicted by GRAIL: only three (exons 2, 4 and 6) of the 14 exons were identified *bona fide* and 7 exons were completely missed (Figure 1B). Yet, the gene is located in a GC rich region and intron-exon boundaries follow the GT-AG rule (Mount, S.M., et al., *Nucleic Acids Research.*, **10**, 459-472 (1982)). Subsequent analysis of the genomic sequence with other gene finding software including GRAIL1a (Uberbacher, E., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 11261-11265 (1991)), Xpound (Thomas, A., & Skolnick, M. H., *IMA J. Math. Appl. Med. Biol.*, **11**, 149-160 (1994)), and Genie (Kulp, D., et al., ISMB-96, St. Louis, MO, AAAI/MIT Press. (http://www-hgc.lbl.gov/projects/genie.html) (1996)) showed that Genie, based on hidden Markov model, performed best for modeling the 3' end of this gene (Figure 1B).

Example 3: APECED-associated mutations found in the APGD1-gene

For mutation screening in APECED patients, all 14 exons were amplified from genomic DNA using primers located in the respective flanking introns (primer sequences and the detailed protocols available at http://chr21.rz-berlin.mpg.de/APECED.html). Five different mutations were identified in the coding region of APGD1 (Table 1). The mutations were monitored in a

control panel of 500 unrelated Finns and 60 unrelated Europeans including 32 CEPH parents. The most common mutation was the "Finnish major mutation" found in 82% of the Finnish patients, all of which have the major disease haplotype (No. 1.1 in Table 1) (Björses, P., et al., *Am. J. Hum. Genet.*, **59**, 8779-886 (1996)). This mutation is a C to T transition at nucleotide 889 in exon 6, changing an Arg into a STOP codon. Among the 500 Finns this mutation was detected in two heterozygotes, indicating a carrier frequency of 1: 250. The same mutation was also found in an Italian and in a German patient, who carried different haplotypes (haplotypes No. 1.2 to 1.4 in Table 1, respectively). Two mutations were found in exon 8. The first one is a duplication of four nucleotides (CCTG) normally found at position 1086 to 1089. The other mutation in this exon is a 13 bp deletion (nucleotides 1085 to 1097) observed in four non-Finnish patients (two British, a Dutch and a German) carrying the same haplotype (No. 2.1 in Table 1). Two other mutations which involve insertion or deletion of a single nucleotide were found in exon 10. The insertion of an A at position 1284 was found in two compound heterozygote Finnish patients having the Finnish major mutation in the other allele. Deletion of a C was found at position 1313 in a French patient homozygous for the disease haplotype (No. 5.1 in Table 1). Mutations and the associated haplotypes are summarized in Figure 3 and Table 1. Northern blot analysis performed on lymphoblast mRNA from patients whose cell lines were available (all Finnish patients) did not show a size difference of the transcript or altered level of expression when compared to control subjects. All the mutations cosegregated with the disease in the respective families and were predicted to result in truncation of the conceptual protein (Table 1). This provides strong evidence that alterations of the APGD1 gene represent the primary cause for the APECED disease.

TABLE 1 Mutations in the APGD1 gene

Mutation	No.	Exon	Nucleotide	Haplotype	No.	Consequence
C_{889}->$T_{Fin\ major}$	1	6	889	(4 3 5 1 2)	1.1	Arg -> STOP, truncated 256 aa protein
				(4 4 7 4 5)	1.2	
				(5 4 2 2 5)	1.3	
				(5 4 5 4 3)	1.4	
4 bp insertion	2	8	1086-1089	(5 3 5 3 3)	2.1	frame shift, truncated 371 aa protein
13 bp deletion	3	8	1085-1097	(4 5 5 4 5)	3.1	frame shift, truncated 372 aa protein
A insertion	4	10	1284	(5 4 3 2 5)	4.1	frame shift, truncated 422 aa protein
C deletion	5	10	1313	(2 10 7 4 5)	5.1	frame shift, truncated 478 aa protein

Table 1 summarizes the mutations and the predicted consequences for the APGD1 putative protein. The APGD1 exons were amplified with intronic primers and initially screened by the SSCP method (Orita, M., et al., *Proc. Natl. Acad. Sci. USA*, **86**, 2766-2770 (1989)). Detected changes were characterized by solid-phase sequencing (Syvänen, A. C., et al., *FEBS Lett.*, **258**, 71-74 (1989)). The haplotypes of the disease chromosomes were constructed from alleles of the markers shown in figure 1A (cen - JA1, D21S1912, PFKL$(CA)_n$, PB1, D21S171 - tel). Haplotype 1.1 is the major haplotype in Finland (Fin major). Haplotypes 1.2 (Italian), 1.3 (German) and 1.4 (German) carry the same mutation as the major Finnish allele. Haplotypes 1.3 and 1.4 are most probably of the same origin since they share the same centromeric alleles. An Italian patient was homozygous for haplotype 2.1 and mutation 2. Haplotype 3.1 was observed as homozygous in one Dutch and in two British patients, and as heterozygous in one German patient. All chromosomes carrying this haplotype have mutation 3. Two Finnish patients were compound heterozygotes for haplotype 4.1 and for mutation 4. Haplotype 5.1 and mutation 5 were found homozygous in a French patient. The detected mutations were monitored against a control panel (see text) by minisequencing (Syvänen, A. C., et al., *Am. J. Hum. Genet.*, **52**, 46-59 (1993)) (mutations 1, 4 and 5) or by size separation of radioactively labeled PCR products on denaturing PAGE (mutations 2 and 3). None of these mutations were detected in a homozygous form in the control subjects. The carrier frequency of the Fin major mutation was observed to be 1:250 in the Finland. This mutation was also found in a heterozygous form in one CEPH parent whereas we did not detect any carriers for the other mutations.

Claims

1. A nucleic acid molecule encoding a (poly)peptide co-segregating in mutated form with Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED) which is

   (a) a nucleic acid molecule comprising a nucleic acid molecule encoding the (poly)peptide having the amino acid sequence of Fig. 2A;

   (b) a nucleic acid molecule comprising the nucleic acid molecule having the nucleotide sequence of Fig. 2A that encodes the amino acid sequence of Fig. 2A;

   (c) a nucleic acid molecule hybridizing to the nucleic acid molecule of (a) or (b); or

   (d) a nucleic acid molecule which is degenerate to the nucleic acid molecule of (c).

2. The nucleic acid molecule of claim 1, wherein said polypeptide has the function of a transcription factor or a transcription-associated factor.

3. The nucleic acid molecule of claim 1 or 2, wherein said polypeptide comprises two double-paired zinc finger motifs.

4. A nucleic acid molecule deviating by at least one mutation from the nucleic acid molecule of claim 1 wherein said mutation co-segregates with APECED and is

   (i) an insertion;

   (ii). a deletion;

   (iii) a substitution; and/or

   (iv) an inversion,

   and wherein said mutation further results in a loss of function or a gain of function of the (poly)peptide as defined in any one of claims 1 to 3.

5. The nucleic acid molecule of claim 4, wherein said insertion, which is a duplication of 4 nucleotides (CCTG) normally found at position 1086-1089, is a 4

nucleotide insertion at the nucleotide position 1085 or 1090 or an insertion of an adenosine at position 1284 of the nucleotide sequence of Fig. 2A.

6. The nucleic acid molecule of claim 4, wherein said deletion is a 13 nucleotide deletion of nucleotides 1085 -1097 or a deletion of the cytosine at position 1313 of the nucleotide sequence of Fig. 2A.

7. The nucleic acid molecule of claim 4, wherein said substitution is a cytosine to thymidine exchange at nucleotide position 889 of the nucleotide sequence of Fig. 2A.

8. The nucleic acid molecule of any one of claims 4 to 7, wherein said loss of function is a loss of macromolecule binding properties.

9. The nucleic acid molecule of any one of claims 4 to 7, wherein said gain of function is involved in molecular interaction.

10. A fragment of the nucleic acid molecule of any one of claims 1 to 9 comprising at least 14 nucleotides.

11. A nucleic acid molecule which is complementary to the nucleic acid molecule of any one of claims 1 to 10.

12. A nucleic acid molecule which is a mammalian homologue of the nucleic acid molecule of any one of claims 1 to 11.

13. The nucleic acid molecule of any one of claims 1 to 12 which is DNA or RNA.

14. A primer pair which hybridizes under stringent conditions to the nucleic acid molecule of any one of claims 1 to 13.

15. A vector comprising the nucleic acid molecule of any one of claim 1 to 14.

16. A host transformed with the vector of claim 15.

17. The host of claim 16 which is a bacterium, a yeast cell, an insect cell, a fungal cell, a mammalian cell, a plant cell, a transgenic animal or a transgenic plant.

18. A process of producing a (poly)peptide as defined in any one of claims 1 to 10, 12 or 13 comprising culturing or raising the host of claim 16 or 17 and isolating said (poly)peptide from said culture or said host.

19. A (poly)peptide encoded by the nucleic acid molecule of any one of claims 1 to 10, 12 or 13 or produced by the process of claim 18.

20. A compound derived from the (poly)peptide of claim 19 and having essential the same three dimensional structure thereof.

21. An antibody that specifically recognizes the (poly)peptide of claim 19 and/or the compound of claim 20.

22. A pharmaceutical composition comprising the nucleic acid molecule of any one of claims 1 to 13, the vector of claim 15, the (poly)peptide of claim 19, the compound of claim 20 and/or the antibody of claim 21.

23. A diagnostic composition comprising the nucleic acid molecule of any one of claims 1 to 13, the primer pair of claim 14, the vector of claim 15, the (poly)peptide of claim 19, the compound of claim 20 and/or the antibody of claim 21.

24. A method for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutation in the nucleic acid molecule of any one of claims 1 to 3.

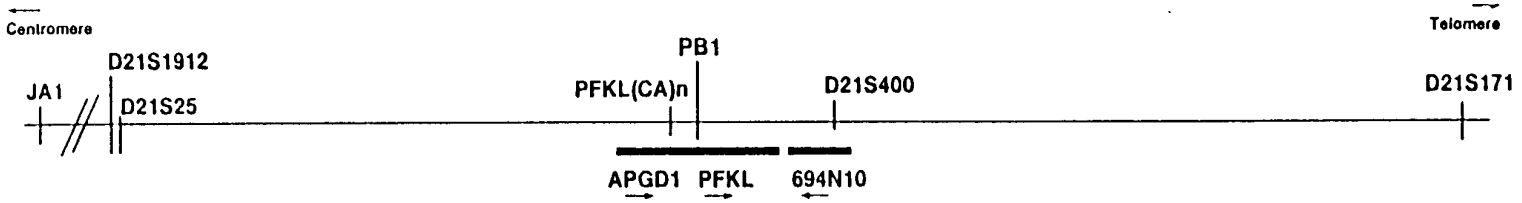
25. A method for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutated form of the polypeptide as defined in any one of claims 1 to 3 in an immuno-assay using the antibody of claim 21.

26. Use of the nucleic acid molecule of any one of claims 1 to 3 or the vector of claim 15 in gene therapy.

Abstract

APECED (Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy) is the only described systemic autoimmune disease with monogenic background, and the first autoimmune disorder localised outside the major histocompatibility complex (MHC) region. We have isolated a novel gene in chromosome 21q22.3 and have identified five different mutations in APECED patients originating from different populations. Thus, we named this gene APGD1 for Autoimmune PolyGlandular Disease type 1. The disease gene encodes a novel 545 amino acid protein of unknown function featuring two PHD double-paired zinc finger motifs previously described in nuclear proteins involved in the regulation of transcription. The identification of the disease gene facilitates direct genetic diagnosis of APECED and provides tools for dissection of the molecular pathogenesis of the disease. Moreover, functional analyses of the novel gene and the corresponding protein will provide insights in basic molecular mechanisms of human autoimmunity.

Figure 1A



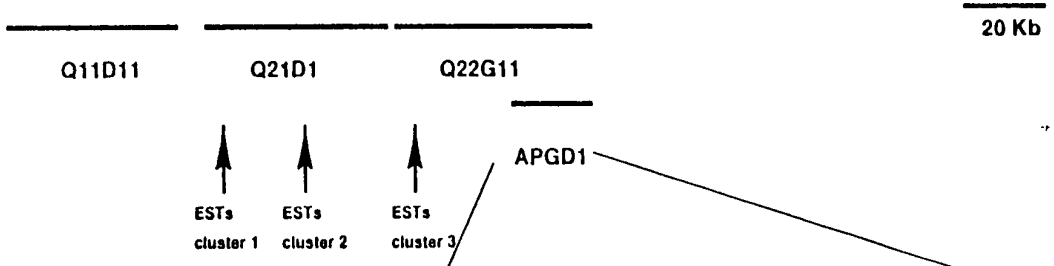
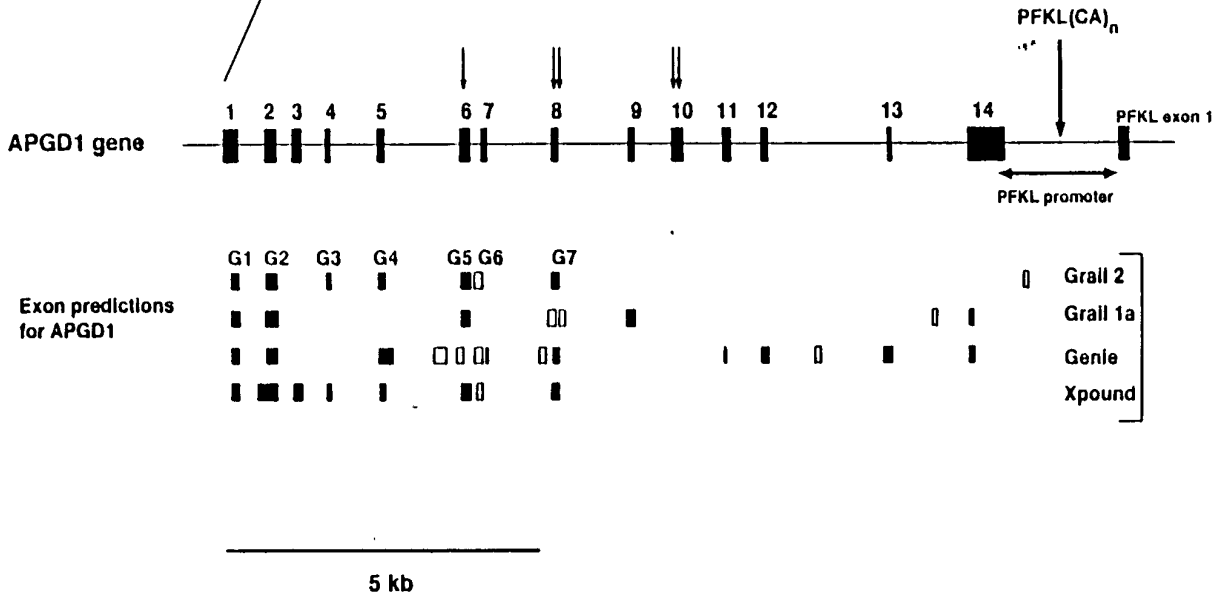


Figure 1B



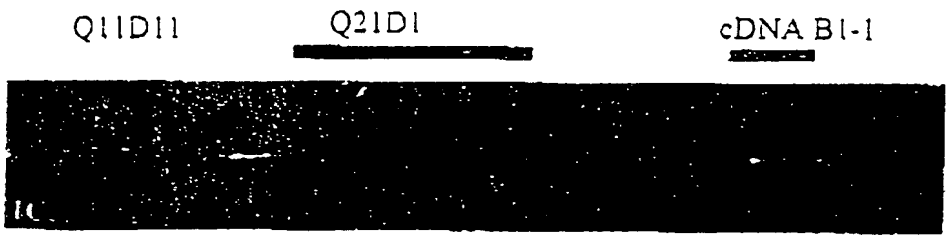


Figure 1C

```
   1 cgggcgcacagccggcgcggaggccccacagccccgccgggacccgaggccaagcgaggg   60

  61 gctgccagtgtcccgggacccaccgcgtccgccccagccccgggtccccgcgcccacccc  120

 121 atggcgacggacgcggcgctacgccggcttctgaggctgcaccgcacggagatcgcggtg  180
   1 M  A  T  D  A  A  L  R  R  L  L  R  L  H  R  T  E  I  A  V   20

 181 gccgtggacagcgccttcccactgctgcacgcgctggctgaccacgacgtggtccccgag  240
  21 A  V  D  S  A  F  P  L  L  H  A  L  A  D  H  D  V  V  P  E   40

 241 gacaagtttcaggagacgcttcatctgaaggaaaaggagggctgcccccaggccttccac  300
  41 D  K  F  Q  E  T  L  H  L  K  E  K  E  G  C  P  Q  A  F  H   60

 301 gccctcctgtcctggctgctgacccaggactccacagccatcctggacttctggagggtg  360
  61 A  L  L  S  W  L  L  T  Q  D  S  T  A  I  L  D  F  W  R  V   80

 361 ctgttcaaggactacaacctggagcgctatggccggctgcagcccatcctggacagcttc  420
  81 L  F  K  D  Y  N  L  E  R  Y  G  R  L  Q  P  I  L  D  S  F  100

 421 cccaaagatgtggacctcagccagccccggaaggggaggaagcccccggccgtccccaag  480
 101 P  K  D  V  D  L  S  Q  P  R  K  G  R  K  P  P  A  V  P  K  120

 481 gctttggtaccgccacccagactccccaccaagaggaaggcctcagaagaggctcgagct  540
 121 A  L  V  P  P  P  R  L  P  T  K  R  K  A  S  E  E  A  R  A  140

 541 gccgcgccagcagccctgactccaaggggcaccgccagcccaggctctcaactgaaggcc  600
 141 A  A  P  A  A  L  T  P  R  G  T  A  S  P  G  S  Q  L  K  A  160

 601 aagccccccaagaagccggagagcagcgcagagcagcagcgccttccactcgggaacggg  660
 161 K  P  P  K  K  P  E  S  S  A  E  Q  Q  R  L  P  L  G  N  G  180

 661 attcagaccatgtcagcttcagtccagagagctgtggccatgtcctccggggacgtcccg  720
 181 I  Q  T  M  S  A  S  V  Q  R  A  V  A  M  S  S  G  D  V  P  200

 721 ggagcccgaggggccgtggaggggatcctcatccagcaggtgtttgagtcaggcggctcc  780
 201 G  A  R  G  A  V  E  G  I  L  I  Q  Q  V  F  E  S  G  G  S  220

 781 aagaagtgcatccaggttggtggggagttctacactcccagcaagttcgaagactccggc  840
 221 K  K  C  I  Q  V  G  G  E  F  Y  T  P  S  K  F  E  D  S  G  240

 841 agtgggaagaacaaggcccgcagcagcagtggcccgaagcctctggttcgagccaaggga  900
 241 S  G  K  N  K  A  R  S  S  S  G  P  K  P  L  V  R  A  K  G  260

 901 gcccagggcgctgcccccggtggaggtgaggctaggctgggccagcagggcagcgttccc  960
 261 A  Q  G  A  A  P  G  G  G  E  A  R  L  G  Q  Q  G  S  V  P  280

 961 gcccctctggccctccccagtgaccccagctccaccagaagaatgaggacgagtgtgcc  1020
 281 A  P  L  A  L  P  S  D  P  Q  L  H  Q  K  N  E  D  E  C  A  300

1021 gtgtgtcgggacggcggggagctcatctgctgtgacggctgccctcgggccttccacctg  1080
 301 V  C  R  D  G  G  E  L  I  C  C  D  G  C  P  R  A  F  H  L  320

1081 gcctgcctgtcccctccgctccgggagatccccagtgggacctggaggtgctccagctgc  1140
 321 A  C  L  S  P  P  L  R  E  I  P  S  G  T  W  R  C  S  S  C  340

1141 ctgcaggcaacagtccaggaggtgcagccccgggcagaggagccccggccccaggagcca  1200
 341 L  Q  A  T  V  Q  E  V  Q  P  R  A  E  E  P  R  P  Q  E  P  360

1201 cccgtggagaccccgctccccccggggcttaggtcggcgggagaggaggtaagaggtcca  1260
 361 P  V  E  T  P  L  P  P  G  L  R  S  A  G  E  E  V  R  G  P  380

1261 cctggggaaccccctagccggcatggacacgactcttgtctacaagcacctgccggctccg  1320
 381 P  G  E  P  L  A  G  M  D  T  T  L  V  Y  K  H  L  P  A  P  400

1321 ccttctgcagccccgctgccagggctggactcctcggccctgcacccccctactgtgtgtg  1380
 401 P  S  A  A  P  L  P  G  L  D  S  S  A  L  H  P  L  L  C  V  420
```

```
1381 ggtcctgagggtcagcagaacctggctcctggtgcgcgttgcggggtgtgcggagatggt  1440
 421 G  P  E  G  Q  Q  N  L  A  P  G  A  R  C  G  V  C  G  D  G   440

1441 acggacgtgctgcggtgtactcactgcgccgctgccttccactggcgctgccacttccca  1500
 441 T  D  V  L  R  C  T  H  C  A  A  A  F  H  W  R  C  H  F  P   460

1501 gccggcacctcccggcccgggacgggcctgcgctgcagatcctgctcaggagacgtgacc  1560
 461 A  G  T  S  R  P  G  T  G  L  R  C  R  S  C  S  G  D  V  T   480

1561 ccagcccctgtggaggggtgctggccccagccccgcccgcctggcccctgggcctgcc  1620
 481 P  A  P  V  E  G  V  L  A  P  S  P  A  R  L  A  P  G  P  A   500

1621 aaggatgacactgccagtcacgagcccgctctgcacagggatgacctggagtcccttctg  1680
 501 K  D  D  T  A  S  H  E  P  A  L  H  R  D  D  L  E  S  L  L   520

1681 agcgagcacaccttcgatggcatcctgcagtgggccatccagagcatggcccgtccggcg  1740
 521 S  E  H  T  F  D  G  I  L  Q  W  A  I  Q  S  M  A  R  P  A   540

1741 gccccctttccctcctgaccccagatggccgggacatgcagctctgatgagagagtgctg  1800
 541 A  P  F  P  S  •                                            546

1801 agaaggacacctccttcctcagtcctggaagccggccggctgggatcaagaaggggacag  1860

1861 cgccacctcttgtcagtgctcggctgtaaacagctctgtgtttctggggacaccagccat  1920

1921 catgtgcctggaaattaaaccctgccccacttctctactctggaagtccccgggagcctc  1980

1981 tccttgcctggtgacctactaaaaatataaaaattagctgggtgtggtggtgggtgcctg  2040

2041 taatcccagctacatgggagcctgaggcatgagaatcacttgaactcgggaggtggaggt  2100

2101 tgcagtgagctgagattgcgccactgcactccagtctggtcggcaagagtgagactccgt  2160

2161 ctcaaaaacaaaacaaaaaaaaccacataacataaaatttatcatctcgaccacttttcagt  2220

2221 tcagtggcattcacatctcatgtaa  2245
```

Figure 2A

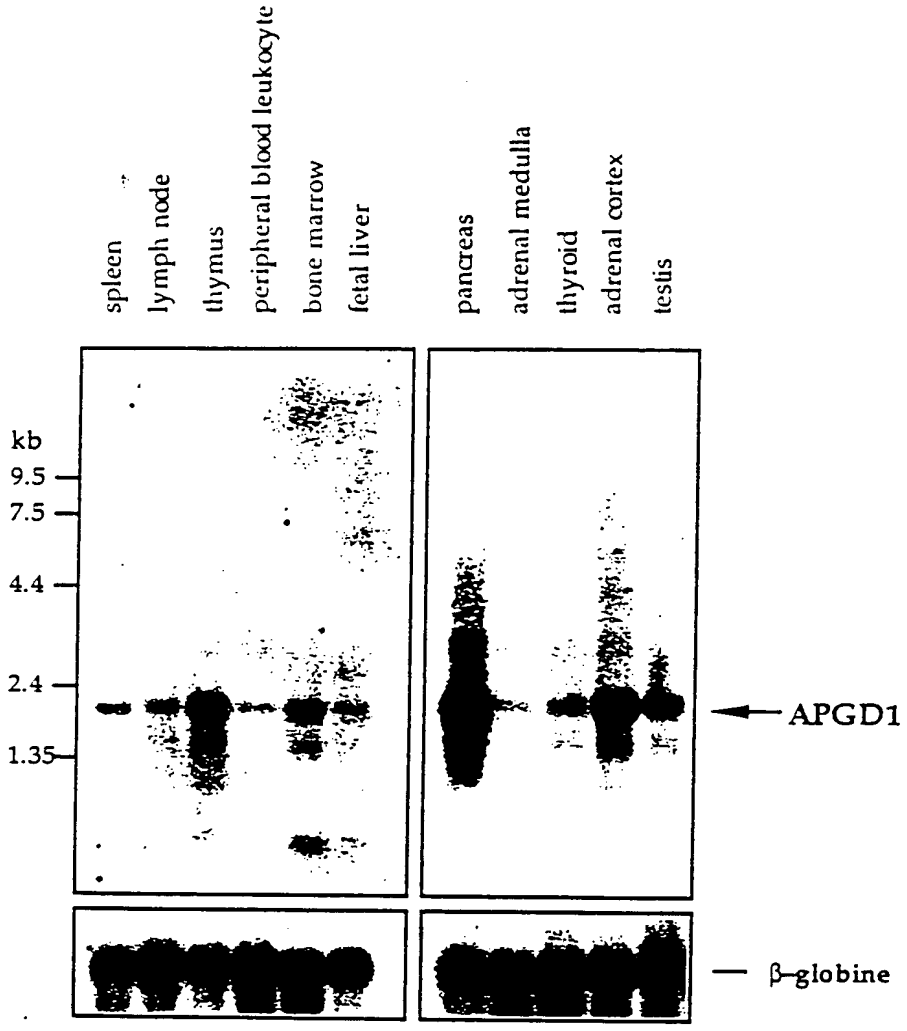

spleen
lymph node
thymus
peripheral blood leukocyte
bone marrow
fetal liver
pancreas
adrenal medulla
thyroid
adrenal cortex
testis
kb
9.5
7.5
4.4
2.4
1.35
APGD1
β-globine

Figure 2B

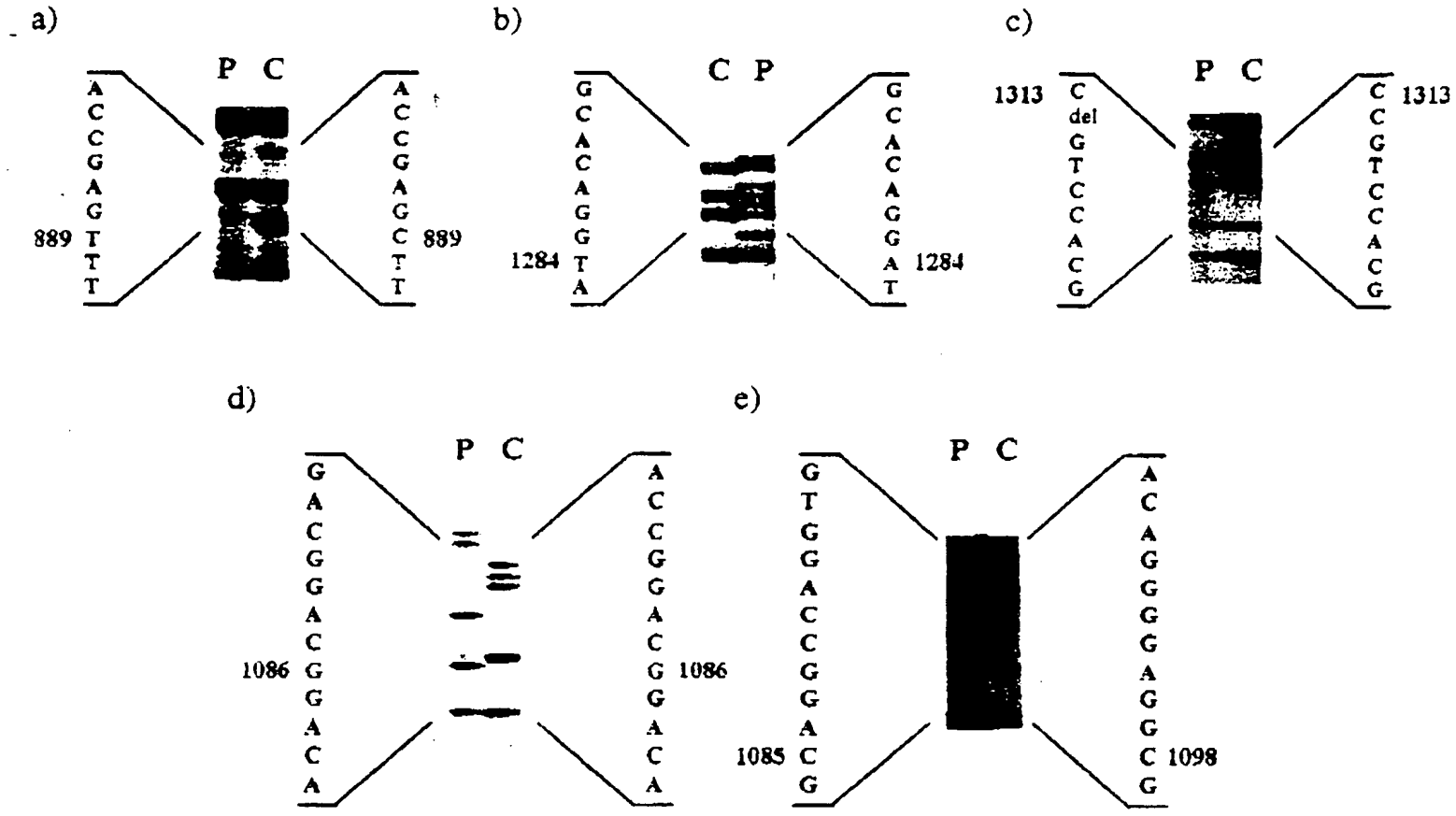

a)
P C
ACCGAGTTT
889
ACCGAGCTT
889
b)
C P
GCACAGGTA
1284
GCACAGGAT
1284
c)
P C
1313
C del GTCCACG
C CGTCCACG
1313
d)
P C
GACGGGACACGGGACA
1086
ACCGGGACGGGACA
1086
e)
P C
GTGGACCGGGACG
1085
ACAGGGGGAGGGC
1098

Figure 3